Considerations for a grid-based Physics Analysis Facility

Dietrich Liko

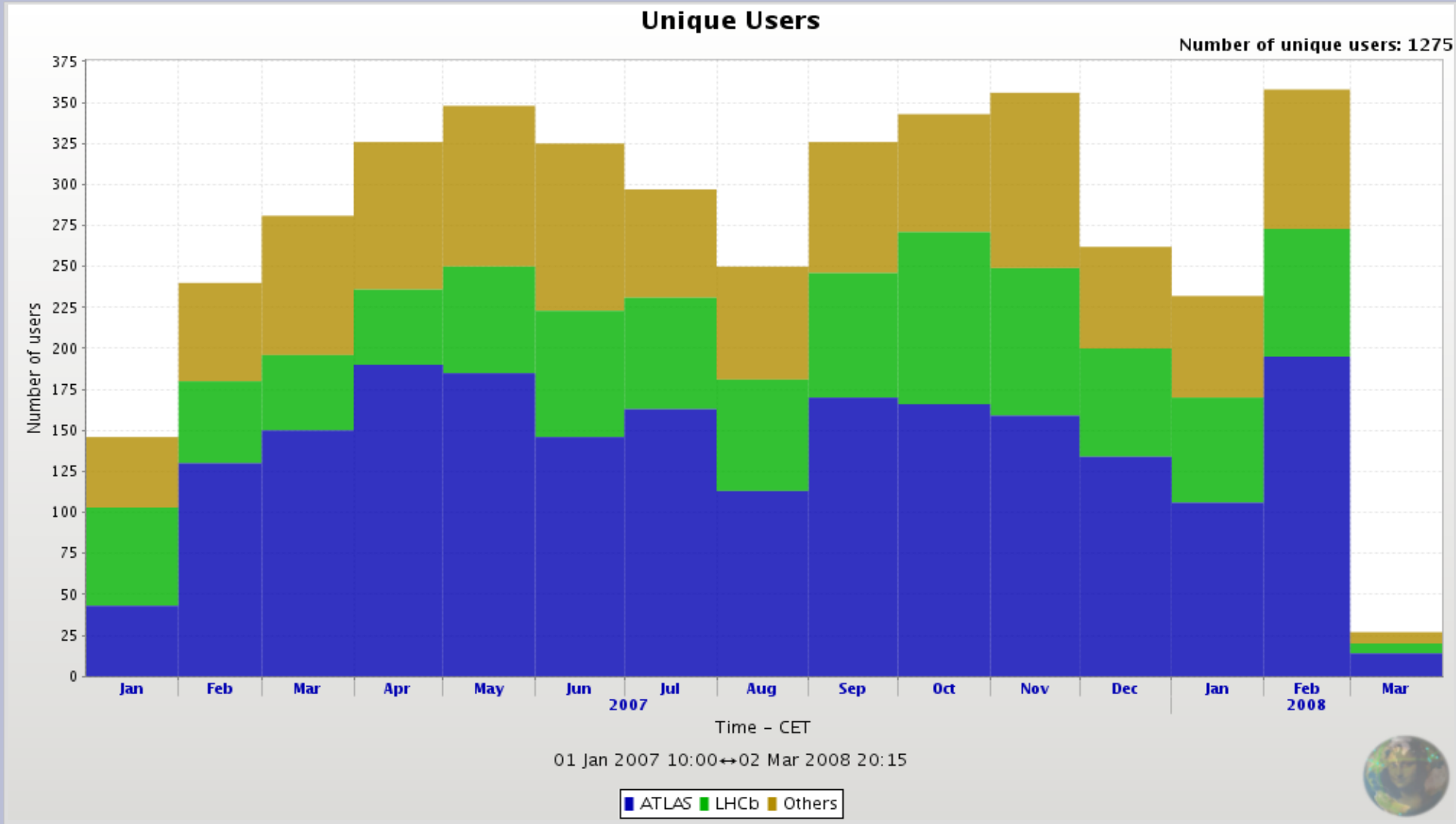
Introduction

- Aim of our grid activities is to enable physicists to do their work
- Latest GANGA developments
  - PANDA
- Tier-3 Taskforce
  - Special workgroup in ATLAS
  - Also concerned with Analysis Facilities
  - Tries to bridge the gap between the grid and local facilities
  - Very relevant for all considerations in Austria

The GANGA system

- User friendly job submission tools
  - Extensible due to pluginsystem
  - Support for several applications
    - Athena, AthenaMC(ATLAS)
    - Gaudi, DaVinci(LHCb)
    - Others …
- Support for several backends
  - LSF, PBS, SGE etc
  - gLiteWMS, Nordugrid, Condor
  - DIRAC, PANDA (under development)

GANGA Usage


Unique Users
Number of unique users: 1275
Number of users
375
350
325
300
275
250
225
200
175
150
125
100
75
50
25
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
Feb
Mar
2007
2008
Time - CET
01 Jan 2007 10:00↔02 Mar 2008 20:15
ATLAS
LHCb
Others

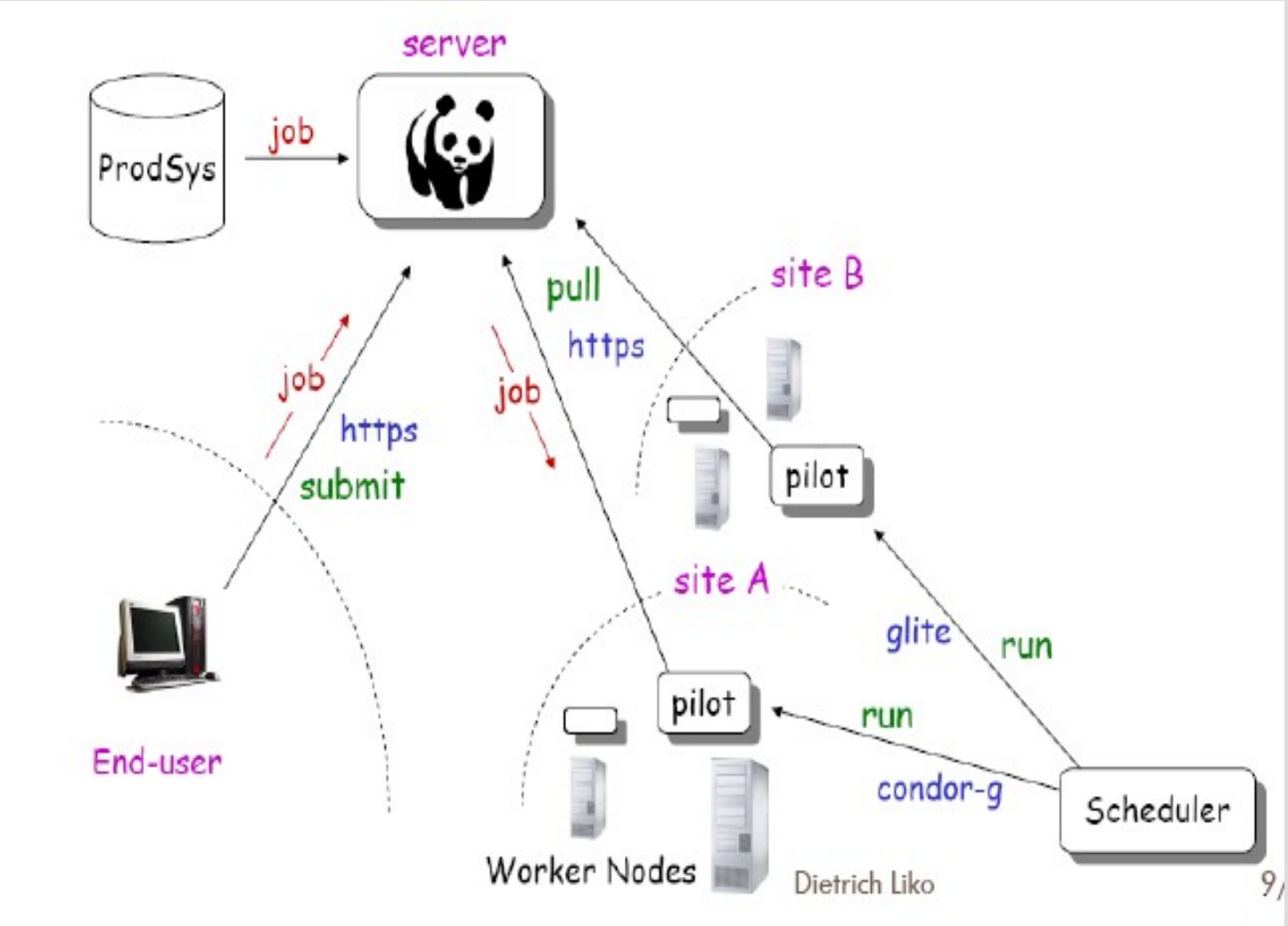
GANGA Plans

- GANGA currently supports 40+ sites
- The PANDA system is being deployed on LCG sites right now
- PANDA will be an excellent backend also for GANGA
- Some collaboration between GANGA and PANDA is being done right now

PANDA for Production

- The PANDA system has been chosen as Production system for ATLAS
- PANDA is a pilot based system
  - Higher efficency
- PANDA incoperates a better support for ATLAS workflows that include data movements

PANDA System

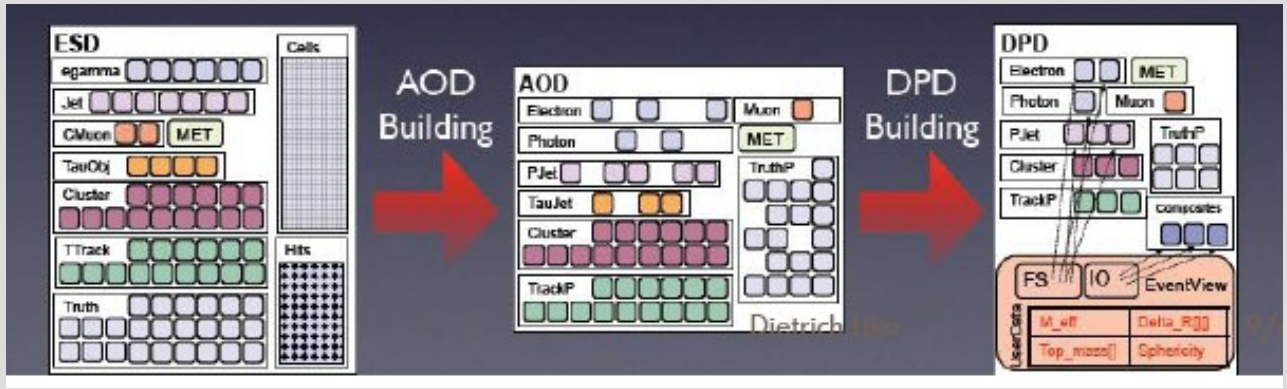


ATLAS Data Formats

- RAW
  - Have to be recorded on permanent storage
  - Only a small fraction can be analyzed directly
- Event Summary Data –ESD
  - Output of the reconstruction
  - Often large; difficult to analyze the full set
- Analysis Object Data –AOD
  - Quantities in particular relevant for physics analysis
  - Main input for analysis, distributed to many sites
- Analysis specific data
  - DPD
  - Thinned AODs and user specific data

ATLAS Analysis Model

- Basic principle: Smaller data can be read faster
  - Skimming-Keep interesting events
  - Thinning-Keep interesting objects in events
  - Slimming-Keep interesting info in objects
  - Reduction-Build higher-level data



Physics Use Cases

- AODs reside at T2, too large to transfer large fraction to T3
- D1PD is 10-20 times smaller, analysis at T3 with ROOT
  - AthenaROOTAccess possible
  - But still would take time, probably better at T2
- Tier 3 will hold D2PD/D3PD
  - 10 TB = 100% of 109 events in 10KB/event D3PDs format
- Rough size estimate
  - 25 cores and 25 TB per analysis (not per user)
  - Can be substantially more!

Analysis Scenaria

- D2PD Production
  - In/out 33/40 KB/event
  - CPU:100 ms/event
- D3PD Production
  - In/out 33/10 KB/event
  - CPU: 50 ms/event
- Plotting
  - In: 1 KB/event
  - CPU: 0

->D2PD	Laptop	Tier 3	Tier 2	Tier 1/2
Cores	1	25	100	1000
1 Hour	>0.01%	0.08%	0.34%	3.36%
Overnight	0.04%	1.01%	4.03%	40.26%
1 Week	0.56%	14.09%	56.37%	All
1 Month	2.42%	60.39%	All	All

->D3PD	Laptop	Tier 3	Tier 2	Tier 1/2
Cores	1	25	100	1000
1 Hour	0.01%	0.17%	0.66%	6.63%
Overnight	0.08%	1.99%	7.96%	79.56%
1 Week	1.11%	27.85%	All	All
1 Month	4.77%	All	All	All

Plotting	Laptop	Tier 3	Tier 2	Tier 1/2
Cores	1	25	100	1000
1 Hour	3.56%	89.02%	All	All
Overnight	42.73%	All	All	All
1 Week	All	All	All	All
1 Month	All	All	All	All

- Assuming perfect hardware/software

Hardware

- Difficult to name vendors and systems…
  - Consult your T1/T2/computing center or neighbored T3 for their HW
  - Combined hardware purchases might lower prices
  - Also check large vendors: Often they have R&D prices and are willing to negotiate
- In the end local considerations

Operating System

- Atlas SW best supported for SL(C)(4)
  - SL4 might be too old for some modern HW
  - Other OS wanted by your administrator
  - Some features need specific OS (Sun Thumper need Solaris for ZFS)
  -
- You can do analysis on other platforms than SLC4. Official support might, however, not be available.

Installation

- Depends on the size of your setup
  - Very few machines: Installation by hand probably most efficient
- Larger clusters: You need installation and maintenance tools.
  - Many HEP and non HEP examples exist
  - Do not forget monitoring

Batch system

- Many are around. If you must choose one:
  - Condor, PBS/Maui/Torque, SGE are freely available and scalable enough for T3
- Choose the one that is best supported in your region
- Example Maui/Torque installation on Wiki
- This combination probably best supported in the EGEE world

PROOF

- Interesting technology, shows good scalability
  - Many Atlas people looking into it:
  - Munich, Madrid, Valencia, BNL, Wisconsin,…
- ALICE relies on it
- Future of PROOF development uncertain (status PROOF WS)
- Summary from Johannes Elmsheuser (more in the Wiki)
  - Dedicate a few nodes to PROOF analysis
  - Use cluster or dCache/xrootd file system
  - Will scale with the 5-10 average user number of an institute
- A sysadmin or computer physicist needs to maintain the cluster

Storage System

- First attemt for a matrix
- Also HEPIX is looking into these issues

Storage System	Local Protocol	Load Balancing	Externally Secure	POSIX Access	Single Namespace	Installation Load	Maint Load	Quotas	Cost
NFS	bad	N	N	Y	N	low	high	Y	$0
Lustre	Y	Y	w/SRM	Y	Y	medium	medium	Y	$0
GPFS	Y	Y	w/SRM	Y	Y	high	medium	Y	$$$
xrootd	Y	Y	w/SRM	mkdir/rmdir do nothing	Y	medium	low	partitions	$0
DPM	Y	Y	Y	special commands	Y	medium-high	low-medium	partitions	$0
dCache	Y	Y	Y	metadata	Y	high	low-medium	partitions	$0

Data import

- Probably done via DDM/DQ2
- SRM endpoint at T3 helpful
- FTS channels: should be available
  - Need some coordination with T1/T2s in your cloud!
  - Users can also write files using lcg-cr directly to their SE from another T2
  - In the future it will be the only place for you to store your own data!
- Probably no export of data necessary, although interesting for regional/national collaborations

Computing Element

- If batch system runs fine: Grid interface is an add-on
- Interesting for regional/national collaborations
- “More work” against “Additional benefits”
- Choose the middleware you know best/ your region supports

Software

- Atlas Software:
  - Automatic installation via Grid possible
  - Manual installation easy
- In any case: Have a shared filesystem with ~100 GB space
  - NFS should be OK
  - AFS also works, Grid installations more difficult
- Grid UI
  - Use the one from CERN AFS
  - Maitain your local copy

Summary

- GANGA is proving an analysis tool for a large community in ATLAS
- Support for users does not stop here
  - Tier-3 or Analysis facility support is necessary
- The Wiki contains more details

Personal summary

- I had a lot of fun supporting ATLAS move to enable the grid for user analysis
- I will move very soon now to CMS and Vienna
- I am looking forward to stay in contact with my friends in ATLAS both at CERN and also in Innsbruck